SOS2 and ACP1 Loci Identified through Large-Scale Exome Chip Analysis Regulate Kidney Development and Function

Two loci identified through large-scale exome chip analysis, *SOS2* and *ACP1*, show evidence for altered kidney development and function

Man Li[1,2]*, Yong Li[3]*, Olivia Weeks[4]*, Vladan Mijatovic[5], Alexander Teumer[6], Jennifer E Huffman[7,8], Gerard Tromp[9,10], Christian Fuchsberger[11,12], Mathias Gorski[13,14], Leo-Pekka Lyytikäinen[15], Teresa Nutile[16], Sanaz Sedaghat[17], Rossella Sorice[16], Adrienne Tin[1], Qiong Yang[18], Tarunveer S Ahluwalia[19,20], Dan E Arking[21], Nathan A Bihlmeyer[21], Carsten A Böger[14], Robert J Carroll[22], Daniel I Chasman[22,4,24], Marilyn C Cornelis[25], Abbas Dehghan[17], Jessica D Faul[26], Mary F Feitosa[27], Giovanni Gambaro[28], Paolo Gasparini[29], Franco Giulianini[23], Iris Heid[13,30], Jinyan Huang[31,32], Medea Imboden[33], Anne U Jackson[12], Janina Jeff[34], Min A Jhun[35], Ronit Katz[36], Annette Kifley[37], Tuomas O Kilpeläinen[19], Ashish Kumar[33], Markku Laakso[38], Ruifang Li-Gao[39], Kurt Lohman[40], Yingchang Lu[34], Reedik Mägi[41], Giovanni Malerba[42], Evelin Mihailov[41], Karen L Mohlke[43], Dennis O Mook-Kanamori[44,45], Antonietta Robino[46], Douglas Ruderfer[47], Erika Salvi[48], Ursula M Schick[49], Christina-Alexandra Schulz[50], Albert V Smith[51,52], Jennifer A Smith[35], Michela Traglia[53], Laura M Yerges-Armstrong[54], Wei Zhao[35], Mark O Goodarzi[55,56], Aldi T Kraja[57], Chunyu Liu[7], Jennifer Wessel[58,59], CHARGE Glycemic-T2D Group, CHARGE Blood Pressure Group, Eric Boerwinkle[60], Ingrid B Borecki[27], Jette Bork-Jensen[19], Erwin P Bottinger[34], Daniele Braga[48], Ivan Brandslund[61], Jennifer A Brody[62], Archie Campbell[8], David J Carey[9], Cramer Christensen[63], Josef Coresh[1], Errol Crook[64], Gary C Curhan[65], Daniele Cusi[48,66], Ian H de Boer[36] , Aiko PJ de Vries[67], Joshua C Denny[22], Olivier Devuyst[68], Albert W Dreisbach[69], Karlhans Endlich[70], Tõnu Esko[41,71,72,21], Oscar H Franco[17], Tibor Fulop[69], Glenn S Gerhard[73], Charlotte Glümer[74], Omri Gottesman[34], Niels Grarup[19], Vilmundur Gudnason[51], Tamara B Harris[75], Caroline Hayward[76,8], Lynne Hocking[77], Albert Hofman[17], Frank B Hu[78], Lise Lotte N Husemoen[74], Rebecca D Jackson[79], Torben Jørgensen[74], Marit E Jørgensen[20], Mika Kähönen[14], Sharon LR Kardia[35], Wolfgang König[80,81,82], Charles Kooperberg[49], Jennifer Kriebel[83,84], Lenore J Launer[75], Torsten Lauritzen[85], Terho Lehtimäki[15], Daniel Levy[7], Pamela Linksted[8], Allan Linneberg[74,86,87], Yongmei Liu[88], Ruth JF Loos[34], Antonio Lupo[89], Christine Meisinger[83], Olle Melander[50], Andres Metspalu[41], Paul Mitchell[37], Matthias Nauck[90,91], Peter Nürnberg[92], Marju Orho-Melander[50], Afshin Parsa[93], Oluf Pedersen[19], Annette Peters[83,84,94], Ulrike Peters[49], Ozren Polasek[95], David Porteous[8], Nicole M Probst-Hensch[33], Bruce M Psaty[62,96,97], Lu Qi[78], Olli T Raitakari[98], Alex P Reiner[49], Rainer Rettig[99], Paul M Ridker[23,100], Fernando Rivadeneira[101], Jacques E Rossouw[102], Frank Schmidt[103], David Siscovick[62], Nicole Soranzo[42], Konstantin Strauch[83,104], Daniela Toniolo[53], Stephen T Turner[105], André G Uitterlinden[101], Sheila Ulivi[46], Dinesh Velayutham[48], Uwe Völker[103,91], Henry Völzke[6,91,106], Melanie Waldenberger[83,84], Jie Jin Wang[37], David R Weir[26], Daniel Witte[107], Helena Kuivaniemi[9,10], Caroline S Fox[7], Nora Franceschini[108], Wolfram Goessling[4,109,110,111,20@], Anna Köttgen[3,1@], Audrey Y Chu[7,23@].

* Indicates joint contribution
@ Indicates joint oversight

Running title: exome-chip association study of eGFR

Word count abstract: 142

Word count text: 2,867

Correspondence:

Audrey Y Chu PhD
NHLBI's Framingham Heart Study
73 Mt Wayte Ave Suite #2
Framingham MA 01702
USA
audrey.chu@nih.gov
+1 508 663 4085

Anna Köttgen MD MPH
University Medical Center Freiburg
Berliner Allee 29
79110 Freiburg
Germany
anna.koettgen@uniklinik-freiburg.de
+49 761 27078050

Wolfram Goessling MD PhD
Genetics Division, Brigham and Women's Hospital and Harvard Medical School
New Research Building, Rm 458
77 Avenue Louis Pasteur
Boston, MA 02115
USA
Wolfram_Goessling@dfci.harvard.edu
+1 617 525 4701

Author Affiliations

1. Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA
2. Division of Nephrology and Department of Human Genetics University of Utah, Salt Lake City, Utah, USA
3. Medical Center – University of Freiburg, Division of Genetic Epidemiology, 79106 Freiburg, Germany

4. Genetics Division, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
5. Department of Life and Reproduction Sciences, University of Verona, 37134 Verona, Italy
6. Institute for Community Medicine, University Medicine Greifswald, 17475 Greifswald, Germany
7. NHLBI's Framingham Heart Study and the Center for Population Studies. Framingham, Massachusetts, USA
8. MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland, UK EH4 2XU
9. The Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, Pennsylvania, USA
10. Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa
11. Center for Biomedicine, European Academy of Bolzano/Bozen (EURAC), affiliated to the University of Lübeck, 39100 Bolzano/Bozen, Italy
12. Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA
13. Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, 93042 Regensburg, Germany
14. Department of Nephrology University Hospital Regensburg, 93042 Regensburg, Germany
15. Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, FI-33101 Tampere, Finland
16. Institute of Genetics and Biophysics "Adriano Buzzati-Traverso", 80131 Napoli, Italy.
17. Department of Epidemiology, Erasmus Medical Center, 3015 GE Rotterdam, The Netherlands
18. Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA
19. Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Universitetsparken 1 ,2100, Copenhagen, Denmark
20. Steno Diabetes Center, Gentofte-2820, Denmark
21. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
22. Vanderbilt University Medical Center Department of Biomedical Informatics, Nashville TN 37203 USA
23. Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA
24. Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
25. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

26. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA
27. Washington University School of Medicine, St. Louis, Missouri, USA
28. Division of Nephrology, Gemelli Foundation University Hospital, Catholic University, 00176 Rome, Italy
29. Institute for Maternal and Child Health - IRCCS “Burlo Garofolo” – University of Trieste, Trieste, Italy
30. Institute of Genetic Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, 85764 Neuherberg, Germany
31. Institute of Hematology, Rui Jin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
32. State Key Laboratory of Medical Genomics, Shanghai, China
33. Swiss Tropical and Public Health Institute, 4002 Basel, Switzerland
34. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1003, New York, New York, USA
35. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA
36. Kidney Research Institute, University of Washington, Seattle, Washington, USA
37. Centre for Vision Research, Department of Ophthalmology and Westmead Millennium Institute for Medical Research, University of Sydney, Westmead, Sydney NSW 2145 Australia
38. Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.
39. Department of Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
40. Biostatistical Sciences; Division of Public Health Sciences; Wake Forest University School of Medicine; Winston-Salem, North Carolina, USA
41. Estonian Genome Center of University of Tartu, 51010, Tartu, Estonia
42. University of Verona, Section of Biology and Genetics, Department of Life and Reproduction Sciences, University of Verona, 37135 Verona, Italy
43. Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA
44. Department of Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
45. Epidemiology Section, Department of BESC, KFSH&RC, Riyadh, Saudi Arabia
46. Institute for Maternal and Child Health - IRCCS “Burlo Garofolo”, Trieste, Italy
47. Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA
48. Department of Health Science, University of Milano, Viale Ortles 22/4, 20139 Milano Italy
49. Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

50. Lund University, Department of Clinical Science, SE----20502 Malmö, Sweden
51. Icelandic Heart Association, Kopavogur, Iceland
52. Faculty of Medicine, University of Iceland, Reykjavik, Iceland
53. San Raffaele Scientific Institute, 20132 Milano, Italy
54. Program in Personalized and Genomic Medicine and Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA
55. Department of Medicine and Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA
56. Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, California, USA
57. Division of Statistical Genomics, Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University, St. Louis, Missouri, USA
58. Department of Epidemiology, Fairbanks School of Public Health, Indianapolis, Indiana, USA
59. Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA
60. The Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas, USA
61. Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark. Vejle Sygehus, Kabbeltoft 25, 7100 Vejle,Denmark
62. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA
63. Department of Internal Medicine and Endocrinology, Vejle Hospital, 7100 Vejle, Denmark
64. Department of Medicine, University of South Alabama College of Medicine
65. Brigham and Women's Hospital, Renal Division and Channing Laboratory, 181 Longwood Avenue, Boston, Massachusetts, USA
66. Division of Nephrology, San Paolo Hospital, Milan, Italy
67. Division of Nephrology, Department of Medicine, Leiden University Medical Center and Leiden University, PO Box 9600, 2300 RC Leiden, the Netherlands
68. University of Zurich, Institute of Physiology, Zurich Center for Integrative Human Physiology (ZIHP), Mechanisms of Inherited Kidney Disorders Group, CH-8057 Zürich, Switzerland
69. University of Mississippi, Department of Medicine, University, Mississippi, USA
70. Department of Anatomy and Cell Biology, University Medicine Greifswald, 17475 Greifswald, Germany
71. Division of Endocrinology, Children's Hospital Boston, Boston, Massachusetts, USA
72. Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA

73. Department of Medical Genetics and Molecular Biochemistry, Temple University School of Medicine, Pennsylvania, USA
74. Research Centre for Prevention and Health, The Capital Region of Denmark, Copenhagen, Denmark, Glostrup University Hospital, 2600, Glostrup Denmark
75. National Institutes on Aging, Maryland, USA
76. MRC Human Genetics, University of Edinburgh, Edinburgh, Scotland, UK EH4 2XU
77. Division of Applied Health Sciences, University of Aberdeen, Aberdeen, Scotland AB25 2ZD UK
78. Harvard School of Public Health, Nutrition, Boston, Massachusetts, USA
79. Ohio State University, Columbia, Ohio, USA
80. University of Ulm Medical Center, Department of Internal Medicine II - Cardiology, Ulm, Germany
81. Deutsches Herzzentrum München, Technische Universität München, Munich, Germany
82. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany
83. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
84. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany
85. Department of General Practice, Aarhus University, 8000, Aarhus, Denmark
86. Department of Clinical Experimental Research, Rigshospitalet, Denmark
87. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
88. Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA
89. University of Verona, Renal Unit, Department of Medicine, University-Hospital of Verona, 37126, Verona, VR, Italy
90. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Greifswald, Germany
91. DZHK (German Centre for Cardiovascular Research), partner site Greifswald
92. Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, Germany
93. Division of Nephrology, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA
94. German Center for Cardiovascular Disease Research (DZHK e.V.), Munich, Germany
95. Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland, UK

96. Departments of Epidemiology and Health Services, University of Washington, Seattle, Washington, USA
97. Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA
98. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, and Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland
99. Institute of Physiology, University Medicine Greifswald, 17475 Greifswald, Germany
100. Cardiovascular Medicine Division, Brigham and Women's Hospital and Harvard Medical Schoo, Boston, Massachusetts, USA
101. Department of Internal Medicine, Erasmus Medical Center, 3015 GE Rotterdam, The Netherlands
102. National Heart, Lung, and Blood Institute (NHLBI), Maryland, USA
103. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, 17475 Greifswald, Germany
104. Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
105. Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota, USA
106. DZD (German Centre for Diabetes Research), partner site Greifswald, Greifswald, Germany
107. Department of Public Health, Aarhus University, 8000, Aarhus, Denmark
108. Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA
109. Harvard Stem Cell Institute, Cambridge, Massachusetts, USA
110. Gastroenterology Division, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
111. Dana-Farber Cancer Institute, Boston, Massachusetts, USA

Abstract

Previous genome-wide association studies (GWAS) have identified >50 common variants associated with kidney function. We performed a two-stage meta-analysis of associations between genotypes from the Illumina exome array and estimated glomerular filtration rate based on serum creatinine (eGFRcrea) among European ancestry participants from the CKDGen Consortium (N_{Stage1}: 111666 and N_{Stage2}: 48343). In single-variant analyses, we identified SNPs at 7 new loci ($P_{Stage1}<3.7x10^{-7}$) associated with eGFRcrea (*PPMJ1, EDEM3, ACP1, SPEG, EYA4, CYP1A1 and ATXN2L*), which were mostly common and annotated as non-synonymous variants. In gene-based analysis, functional rare variants in three genes, including the novel *SOS2*, were associated with eGFRcrea ($P_{SKAT}=5.4x10^{-8}$). Experimental follow-up in zebrafish embryos revealed kidney development alterations, including glomerular and renal tubule changes in the embryonic kidney of *acp1*- and *sos2*-knockdowns. These developmental abnormalities are coupled with changes in kidney function, including altered blood clearance and heightened edema prevalence. This study expands the number of loci associated with kidney function and identifies novel genes with potential roles in kidney formation.

Introduction

Chronic kidney disease (CKD) is considered a complex phenotype with a genetic predisposition[1]. Previous GWAS have successfully identified multiple common genetic risk variants associated with the CKD-defining measures of estimated glomerular filtration rate (eGFR) and urinary albumin-to-creatinine ratio (UACR).[2-5] Together, these variants explain only a small proportion of the variation in eGFR and UACR.[6] To comprehensively interrogate protein-coding regions and assess the effects of rare variants (minor allele frequency [MAF]<1%), we carried out a two-stage meta-analysis of the association between eGFRcrea and variants genotyped on the Illumina HumanExome chip [http://genome.sph.umich.edu/wiki/Exome_Chip_Design] among 111666 European ancestry (EA) participants from the CKDGen Consortium and assessed the role of genes significantly associated with eGFRcrea in kidney development using embryonic zebrafish models. In secondary analyses, we examined associations with eGFRcrea stratified by diabetes status, and in a smaller subset of EA participants, we also tested estimated glomerular filtration rate based on cystatin C (eGFRcys) and UACR. An additional 9624 participants of African ancestry (AA) were also used in an independent exome-chip discovery meta-analysis.

Results

Up to 120357 participants from 27 studies of EA and up to 11386 participants from 7 studies of AA were included in Stage 1 analyses for

eGFRcrea, eGFRcys, or UACR. An additional 48343 participants from 12 studies of EA were included in Stage 2 analysis of eGFRcrea. All participants provided informed consent and each of the studies were approved by its governing ethics committee or Institutional Review Boards. Sample characteristics and genotyping information for each study are summarized in Supplementary Tables 1 and 2.

In Stage 1 single-variant EA analyses, we identified 33 loci associated with eGFRcrea (Table 1 and Supplementary Table 3) that met our *a priori* chip-wide significance threshold of $P<3.7x10^{-7}$(0.05/134299 variants). Of these, 8 had not been identified in association with eGFRcrea from previous GWAS analyses; 6 were missense variants - rs34611728 (*PPM1J*), rs78444298 (*EDEM3*), rs11553746 (*ACP1*), rs2307394 (*ORC4*), rs55760516 (*SPEG*), and rs9493627 (*EYA4*) - and 2 were GWAS tag SNPs included on the exome chip due to prior associations in the NHGRI GWAS Catalog for caffeine/coffee intake[7, 8] - rs2472297 (intergenic, near *CYP1A1*) and inflammatory bowel disease[9] - rs8049439 (intronic, *ATXN2L*); all were common variants (MAF>1%). A Manhattan plot displaying all 33 chip-wide significant loci is shown in Figure. 1. Quantile-quantile (QQ) plots indicated no inflation of the overall p-value distribution (Supplementary Figure. 1). Regional association plots show the associations of other variants within the 500 Mb region of the index variant of the 8 newly identified loci in Supplementary Figure. 2.

In Stage 2 analyses, we followed up the 8 newly identified eGFRcrea loci meeting our significance threshold among an additional 48343 EA participants

from 12 studies (sample characteristics and genotyping information are summarized in Supplementary Tables 2 and 4). All loci but *ORC4* met criteria for replication (a direction of effect consistent with Stage 1 analysis, a $P_{1\text{-sided}}<0.05$ from Stage 2 analysis, and $P<3.7\times10^{-7}$ from a combined Stage 1 and Stage 2 analysis, Table 1). Because diabetes mellitus is a major risk factor for CKD, we assessed the genetic associations at these loci in the presence or absence of diabetes to obtain additional insights into potential mechanistic pathways. In total, 11040 and 94677 participants were included in the diabetes stratified and non-diabetes stratified analyses, respectively. In analyses stratified by diabetes status, the beta-coefficients were directionally consistent and of similar magnitude between the strata for 6 out of 7 newly identified loci (Supplementary Table 5).

To identify additional novel loci associated with alternative measures of kidney function, we tested the association of single variants with eGFRcys (N=32861 EA participants) and UACR (N=31164 EA participants). All observed associations achieving the significance threshold ($P<3.7\times10^{-7}$; Supplementary Table 8) have been previously reported in association with eGFRcrea, eGFRcys or UACR.[3, 4, 6]

To further investigate the role of rare variants in kidney function, we performed gene-based tests for 9990 autosomal genes that contained at least 2 non-synonymous or splice-site variants with MAF<1%. No evidence of inflation was observed in QQ plots for gene-based tests (Supplementary Figure. 3). Three genes, *SOS2*, *SLC47A1* and *LRP2*, met the experiment-wide threshold for

gene-based significance ($P<2.5x10^{-6}$, 0.05/19922 tests [9961 genes x 2 tests], Table 2). The association for *SLC47A1*, a renal solute transporter, was driven by the presence of a single variant, rs111653425, with MAF ~ 1% (Supplementary Table 6). Common variants at *SLC47A1* and *LRP2* have been previously reported in association with eGFRcrea,[4, 6] thus implicating both rare and common variants at both loci.[4, 6] Conversely, the association with *SOS2* was novel ($P_{SKAT}=5.38x10^{-8}$; $P_{T1}=3.25x10^{-6}$). No genes reached the threshold for chip-wide significance in gene-based associations for eGFRcrea stratified by diabetes status, eGFRcys, or UACR.

To identify genes that may play a role during kidney development, we used knockdown zebrafish embryos generated by injecting morpholino oligonucleotides (MOs) into single-cell stage embryos. Morpholinos are a commonly used tool in zebrafish screens because they enable an efficient identification of genes that may have a role in developmental processes. The morpholinos targeted genes with non-synonymous variants (*ppmj1*, *acp1*, *eya4*, *speg*, and *edem3*) and the novel gene-based finding, *sos2* (Supplemental Figure 5A). General defects in the pronephros or embryonic kidney structure (marked by expanded *pax2a* expression) were observed in *acp1* ATG MO- and *sos2* ATG MO-injected embryos compared to controls (Figure. 2, panels A, E and I; P<0.0001 for both). Both *acp1* ATG- and *sos2* ATG-knock-downs showed elongated proximal tubules (increased *slc20a1a* expression) compared to the control group (Figure. 2, panels B, F and J; mean difference in proximal tubule length: *sos2*=81.7 μm; P<0.0001 and *acp1*= 74.7 μm; P<0.0001, panel M). The

115 increase in proximal tubule length was not a consequence of increased embryo
116 length, as both *sos2* ATG and *acp1* ATG-morphants had significantly reduced
117 body length relative to controls (Supplementary Figure 5B). Additionally, *acp1*
118 ATG-knockdowns showed shorter distal tubule length (*slc12a3* expression),
119 which may be a consequence of reduced body length (Figure. 2, panels C, G and
120 K; $P<0.0001$). No abnormalities were observed for podocytes (*wt1a* expression)
121 for *sos2* or *acp1* compared to controls (Figure. 2, panels D, H and L; $P>0.05$).
122 Due to the potential off target effects of morpholinos, these developmental
123 findings were validated with secondary splice-site morpholinos designed to target
124 the *sos2* and *acp1* pre-mRNA (Supplementary Figure 5A). The *sos2* and *acp1*
125 splice morpholino-injected embryos had significantly increased proximal tubule
126 length relative to controls (Supplementary Figure 5C-E). Furthermore, the *acp1*
127 splice-site morpholino injected embryos had shortened distal tubules (*slc12a3*
128 expression) consistent with the phenotype induced by the ATG morpholino
129 (Suppementary Figure 5F). The reproducibility of the proximal tubule
130 developmental defects with ATG and splice-site morpholinos adds confidence to
131 the specificity of the morpholino-induced tubule phenotype.
132 Follow-up *in situ* hybridization experiments to determine expression
133 patterns of *sos2* and *acp1* during zebrafish development did not reveal kidney-
134 specific expression of *sos2* or *acp1;* however, *sos2* and *acp1* were broadly
135 expressed throughout embryogenesis at key stages of kidney development and
136 may be acting to control kidney development (Supplementary Figure 6). In
137 humans, both *sos2* and *acp1* protein are detected in adult renal tubules[10]. No

significant developmental abnormalities were observed among MO knockdowns for the remaining genes (Supplementary Figure 6C). These findings suggest that both *SOS2* and *ACP1* may influence embryonic renal development, and that genetic influences on kidney development may contribute to variation in kidney function.

We next sought to determine whether *sos2* and *acp1*-mediated developmental alterations led to abnormalities in kidney function. In zebrafish, edema is a common sign of kidney failure [11]. We first performed an edema prevalence study and identified incidence rates of pericardial and global edema in *sos2* and *acp1* morphant larva. Both *sos2* and *acp1* ATG morpholino injected embryos had a heightened incidence of pericardial edema beginning at 72 hpf (Figure 3A-C). The *sos2* morphants developed severe global edema by 120 hpf, while the *acp1* morphants presented with only pericardial edema (Figure 3A-C). Both *sos2* and *acp1* morphants showed indications of embryonic lethality by 120 – 144 hpf (Figure 3A-C).

An additional metric for kidney function is the assessment of glomerular filtration and fluid flow by fluorescent dextran clearance [12]. Control, *sos2,* or *acp1* morphant embryos were injected with equal volumes of rhodamine-labeled 70 kDA MW dextran in the cardiac sinus venosus at 72 hpf. Dextran clearance rate was assessed by the quantification of rhodamine fluorescence intensity in a standardized area of the cardiac region at 2, 24, and 48 hours post injection (hpi). *sos2* and *acp1* morphant embryos exhibited decreased dextran clearance at both 24 and 48 hpi relative to controls (Figure 3D-I, M). Furthermore, renal tubules

161 marked by fluorescent dextran in *sos2* and *acp1* morphant larva displayed an
162 abnormal morphology (Figure 3J-L). Specifically, the proximal convoluted tubules
163 were reduced in size and lacked coiling depth. Failed clearance of 70 kDa MW
164 dextran and abnormal tubular structure suggests that morphants may have
165 defects in tubular fluid flow and glomerular filtration.
166 We also evaluated heart rate in the *sos2* and *acp1* morphants relative to
167 controls because compromised cardiovascular function could contribute to
168 defects in dextran clearance. At 96 hours post fertilization (24 hours after dextran
169 injection), the *sos2* (145.5 ± 3.854 bpm; p = 0.0028) and *acp1* (151.2 ± 2.653
170 bpm; p < 0.0001) morphants had an elevated mean heart rate relative to controls
171 (127.8 ± 3.353 bpm) (Supplementary Figure 2G). While altered kidney
172 development, abnormal tubular structure, edema, and decreased dextran
173 clearance all support the conclusion that both *sos2* and *acp1* are regulators of
174 kidney development and function, it remains possible that altered hemodynamics
175 contribute to the edema and dextran clearance phenotypes observed.
176 In separate single variant analyses among 9624 AA participants, we
177 identified 3 loci in association with eGFRcrea at chip-wide significance
178 (Supplementary Table 8, $P<3.7x10^{-7}$). These variants were rare and the limited
179 availability of AA cohorts prevented replication of these findings. The *APOL1* G1
180 variant, rs73885319, that were known risk factor for kidney function decline and
181 end-stage renal disease in African Americans, was included on the exom array.
182 However, it is not associated with eGFRcrea in the AA participants here (P=0.70).

To investigate if the newly identified eGFRcrea loci were also associated with diabetes mellitus and arterial hypertension, major risk factors for CKD, we tested for associations between the 7 validated eGFRcrea loci with blood pressure and type 2 diabetes among EA participants in collaborations with the CHARGE Blood Pressure[13] and CHARGE Glycemia-Type 2 Diabetes[14] Working Groups. Consistent with prior observations, the majority of variants were not associated with blood pressure traits or diabetes (Supplementary Table 9). The exception was rs2472297 at the *CYP1A1* locus, a GWAS tag SNP previously associated with coffee/caffeine intake[7, 8]. This SNP was also associated with systolic and diastolic blood pressure ($P=7.4x10^{-7}$, and $4.6x10^{-11}$, respectively).[13]

Discussion

Our main findings are four-fold. First, we identified and validated 7 loci associated with eGFRcrea through genotyping on the exome-chip among 160009 participants of EA in a two-stage study design. Second, the majority of the newly uncovered associations were for common variants with modest effect sizes, which argues against the presence of rare protein-coding variants with large effect sizes represented on the exome chip. Third, we identified one novel association for *SOS2* through gene-based testing. Fourth, we demonstrated altered kidney development and function in zebrafish *sos2* and *acp1* knockdowns.

Our study emphasizes the continued success of efforts combining population-based genetics and model organisms to identify genes underlying kidney function.[4, 15] The zebrafish knockdowns provide a systematic model to

examine the consequences of gene perturbations in the embryonic renal system of the fish. Our zebrafish morpholino experiments revealed a potential role for *sos2* and *acp1* in kidney development and function. Mutations in the *SOS* gene family (*SOS1* and *SOS2*) lead to Noonan syndrome, a congenital RASopathy (syndromes caused by germline mutations controlling signal transduction pathways) that can feature mild kidney dysfunction[16] and renal anomalies.[17, 18] The *SOS2* protein is expressed in the glomeruli and tubules of kidneys from adult humans[10]. Together, these findings provide further evidence for a potential role for *SOS* proteins in kidney development and function. There are no prior reports of an association between kidney function and *ACP1*, a gene that encodes for an acid phosphatase involved in the immune response and found in erythrocytes.[19] Our observations of kidney abnormalities in *sos2*- and *acp1*-knockdowns provide genes and target tissues for prioritization in future studies of more extensive functional follow-up, diagnostic screening and potentially drug development.

Although morpholinos are an efficient tool for the rapid evaluation of GWAS hits, they have the potential for off-target effects, and morphant phenotypes are not always recapitulated in genetic mutant models. In this study, we evaluated *sos2* and *acp1* kidney development phenotypes using two independent morpholinos, which we believe adds confidence in the specificity of our phenotypes. Furthermore, we provided evidence that kidney developmental changes correlate with edema and reduced dextran clearance from the blood. We believe that our morpholino screen has allowed us to clarify promising candidates for further study. Simultaneously, we acknowledge that future studies

in genetic mutants will enhance and clarify these findings. Since genetic knockout techniques do not necessarily recapitulate exact features of the identified human variants, future experiments are also needed to evaluate the impact of specific variants on kidney development and function.

The majority of identified novel variants were common. The strongest single-variant association with eGFRcrea to date is the common variant rs13329952 at the *UMOD* locus with an effect size of 0.016 ln(ml/min/1.73 m^2) (MAF=0.19).[15] Given our large discovery sample, our study was adequately powered (>80%) to detect effect sizes of 0.11 to 0.008 ln(ml/min/1.73 m^2) for very rare variants to more common variants (0.0005>MAF>0.10) (details of the power calculation can be found in Supplementary Methods). While the selection of non-synonymous content was expected to enrich for functional variants with large effect sizes, a possible reason for the lack of these findings might be the design-based, limited coverage of rare variants on the exome chip. The exome chip was primarily designed to assess nonsynonymous rare variants that had been observed among ~12,000 sequenced participants that were not selected based on kidney disease (http://genome.sph.umich.edu/wiki/Exome_Chip_Design), and thus, we would not expect kidney disease causing mutations to be well represented on the exome chip. While the convenience and low cost of a chip-based genotyping array focused on exonic variants were influential in facilitating a large number of cohorts to participate needed for a well-powered study, the limited coverage of the exome impacted the ability of both single-variant and gene-based tests to assess all exonic variation in association with kidney

function. Thus, we cannot rule out *bona fide* rare variant associations with eGFRcrea. Large-scale whole exome or whole genome sequencing will be able to adequately address the unresolved issue of the contribution of rare variants to the variation of kidney function in the general population. Due to limited sample size in the diabetes stratum compared to the non-diabetes stratum, we did not have adequate power to assess unique genetic associations in the diabetes stratified analysis. We did not implement a random effects meta-analysis as an alternative screening procedures for novel loci because the between-study heterogeneity was small. The analysis module that we used cannot correctly account for the different allele dosage between women and men and therefore we were unable to implement association analysis for chromosome X variants. While our study supports the observation that gene-based analyses aggregating rare non-synonymous variants are able to identify new loci in association with common complex phenotypes[14, 20-22] and can additionally uncover new rare missense variants within known loci, the number of loci identified through gene-based methods remains a small minority of the findings compared to single-variant results. Finally, the zebrafish knockdowns helped us to screen novel loci that appear to have a role in kidney development or function. However, extrapolation of the relationship between our novel loci among patients with more advanced CKD remains to be determined; follow up studies are needed to assess the association of these loci with end-stage renal disease and incident CKD.

In summary, we identified 8 novel loci (7 common single variants and 1 gene with multiple rare coding variants) associated with kidney function. Functional experiments in zebrafish highlighted potential roles for *SOS2* and *ACP1* in embryonic kidney development. Future whole exome and whole genome sequencing studies will be needed to assess the full spectrum of rare genetic variants on kidney function.

Concise Methods

Study Participants

Across all traits analyzed, a total of 120357 participants from 27 studies of European ancestry (EA) were included in Stage 1 of this study. An additional 48343 participants from 12 studies of EA were included in Stage 2. A total of 11386 participants from 7 studies of African ancestry (AA) were also included in separate analyses. Study-specific characteristics are summarized in Supplementary Table S1, S2 and S4. All participants provided informed consent and each study was approved by its governing ethics committee or Institutional Review Board.

Phenotype Definitions

Serum creatinine was measured in each study as described in the Study-Specific Methods section in the Supplemental Material, and calibrated to the National Health and Nutrition Examination Study (NHANES) data to account for between-laboratory variation.[23, 24] eGFR based on serum creatinine (eGFRcrea)

297 was estimated using the four-variable MDRD Study Equation.[25] Cystatin C, an
298 alternative biomarker of kidney function, was measured in a sub-set of
299 participating studies. eGFR based on cystatin C (eGFRcys) was estimated as
300 76.7x(serum cystatin C).[25] All eGFRcrea and eGFRcys values <15
301 ml/min/1.73m^2 were set to 15, and those >200 ml/min/1.73m^2 were set to 200 to
302 avoid undue influence from outliers. UACR was defined as urinary albumin
303 (mg/L) / urinary creatinine (mg/dl)]*100. All analyzed traits (eGFRcrea, eGFRcys,
304 and UACR) were natural log (ln)-transformed. Diabetes was defined as fasting
305 glucose ≥126 mg/dl, pharmacologic treatment for diabetes, or by self-report.
306 Hypertension was defined as systolic blood pressure ≥140 mmHg or diastolic
307 blood pressure ≥90 mmHg or pharmacologic treatment for hypertension.[15]
308
309 *Genotypes*
310 Genotyping was conducted in each study using the Illumina Human
311 Exome BeadChip [http://genome.sph.umich.edu/wiki/Exome_Chip_Design.].
312 This genotyping array containing 247870 markers focuses on exonic variants
313 discovered through exome sequencing of ~12000 individuals. Illumina's
314 GenTrain version 2.0 clustering algorithm in GenomeStudio, zCall,[26] or a
315 combination of both procedures were used to call genotypes. To improve
316 genotype calling of low frequency and rare variants, genotypes from 8 of the
317 contributing cohorts (ARIC, AGES, CHS, FHS, RS, Health ABC, FamHS, and
318 JHS) were jointly clustered and called via the CHARGE Consortium algorithm.[27]
319 Other participating cohorts were called individually. Among them, CROATIA-

Korcula and GS applied the cluster file from the CHARGE Consortium for genotype calling. Details regarding genotyping and quality control within each study are summarized in Supplementary Table S2.

Statistical methods for Stage 1

By following a centralized analysis plan, each study performed two sets of analyses: single-variant analysis and gene-based analysis. The primary meta-analyses were focused on the EA population and a secondary set of meta-analyses were focused on the AA population due to a substantially smaller sample size. Where not specified otherwise, R software was used for data management, statistical analyses and graphing.[28]

Single-variant analysis

Each study performed association analyses of the following phenotypes and models: 1) ln-transformed eGFRcrea, 2) ln-transformed eGFRcys, 3) ln-transformed UACR and 4) ln-transformed eGFRcrea stratified by diabetes status. These association analyses were based on linear regression models adjusting for age, sex, study site (if applicable), family structure (if applicable), and the first 10 principal components to control for population stratification. All analyses were performed assuming an additive genetic effect and all analyses were stratified by ancestry.

For single-variant meta-analysis, study-specific results were combined for each trait in a fixed-effects model using METAL.[29] We restricted single-variant

meta-analyses to 1) autosomal variants; 2) polymorphic variants; 3) variants existing in the joint calling effort within the CHARGE consortium;[27] and 4) with minor allele count (MAC) ≥20 across all cohorts in each ancestry group. Bonferroni correction for the number of variants tested within each ancestry was used to set the significance threshold for each analysis (chip-wide significance), corresponding to $P<3.7\times10^{-7}$ (0.05/134299 variants) for EA analysis and $P<5.5\times10^{-7}$ (0.05/91187 variants) for AA analysis.

Gene-based analysis

Single-variant analysis methods have limited power to detect association for rare variants. Recently developed gene-based methods, where defined variants contained in a gene region are collapsed into one unit for analysis, provide additional statistical power[30] to investigate the role of rare variants on kidney function traits.

For gene-based meta-analysis, study-specific results were combined using the seqMeta package for R.[31] The same phenotypes and covariates were used as for single-variant testing and analyses were again stratified by ancestry. We used two gene-based tests for aggregated analysis of rare SNVs: 1) T1,[30] which is more powerful when all variants within the gene region affect the phenotype in the same direction; and 2) the sequence kernel association test (SKAT),[32] which allows for bidirectional effects and is more powerful when there are both protective and deleterious variants within the same gene. Both gene-based tests were restricted to variants with MAF<1% and variants likely to exert a

major effect on the gene product (stop gain/loss, nonsynonymous, or splice-site variants based on annotation with dbNSFP (v.2.0).[33] Genes containing at least two variants with a cumulative MAC ≥20[20] were included in the analysis. In total, we tested 9990 autosomal genes meeting our established thresholds and filters. Bonferroni correction for the number of genes and tests performed was used to set the significance threshold for the gene-based analysis corresponding to $P<2.5x10^{-6}$ (0.05/19922 tests [9961 genes x 2 tests]) in EA analysis and $P<2x10^{-6}$ (25378 tests [12689 genes x 2 tests]) in AA analysis.

Statistical methods for Stage 2

Chip-wide significant results from single-variant meta-analyses were brought forward for testing in Stage 2 where each new study followed the same methodology as Stage 1. Details regarding the genotyping and population characteristics of each cohort can be found in Supplementary Tables S2 and S4.

Study-specific results from Stage 2 cohorts were combined using the same meta-analysis approach and software as in Stage 1 (fixed-effects model in METAL[29]). Criteria for validation were: 1) a direction of effect consistent with Stage 1 analysis, and 2) a one-sided p-value<0.05 from Stage 2 analysis and 3) $P<3.7x10^{-7}$ from combined Stage 1 and Stage 2 analysis.

The percentage of phenotypic variance explained by each novel locus was estimated as $R^2=\beta^2 var(SNP)/var(ln(eGFRcrea))$, where β^2 is the estimated effect of the SNP on ln(eGFRcrea), and $var(SNP)=2*MAF_{SNP}*(1-MAF_{SNP})$. Var(ln(eGFRcrea)) was estimated in the ARIC study. All loci were assumed to

389 have independent effects on the phenotype.

390

391 *Associations with other traits*

392 To examine potential associations of the novel eGFR-associated variants
393 with other correlated traits and conditions, we performed external look ups for
394 systolic and diastolic blood pressure in collaboration with the CHARGE Blood
395 Pressure Working Group (N=145872)[13] and for type 2 diabetes (T2D) in
396 collaboration with the CHARGE Glycemia-T2D Working Group (N=10240 T2D
397 cases and 63105 controls)[20] among EA participants.

398

399 *NHGRI GWAS Catalog and PubMed Queries*

400 For all newly identified and validated variants, we interrogated the NHGRI
401 GWAS Catalog [https://www.ebi.ac.uk/gwas/;10/12/15] for the lead SNP at each
402 locus and for SNPs in linkage disequilibrium with the lead SNP (within 1 Mbp and
403 r^2>0.5 from 1000 Genomes Pilot in CEU;
404 http://www.broadinstitute.org/mpg/snap/) to assess association with other traits
405 (see Supplementary Table S11 for full listing of GWAS Catalog associations).
406 Additionally, for each locus we searched PubMed
407 [http://www.ncbi.nlm.nih.gov/pubmed;10/12/15] for publications on kidney/renal
408 function or chronic kidney disease.

409

410 *Functional Studies in Zebrafish*

To investigate a potential role of the newly identified genes during kidney development, we assessed the functional consequences of gene knockdown in zebrafish embryos. We used antisense morpholino oligonucleotide (MO) technology to knock down genes identified based on validated associations for nonsynonymous variants and for novel gene-based loci. Two independent MO probes were used for each gene. Zebrafish were maintained in accordance with established IACUC protocols.

Morpholinos (Gene Tools) were designed against zebrafish target genes. Morpholino sequences are as follows: *sos2* (5' GCACCGGGAACAACCACACAACTTT 3'), *sos2* (exon 2) (5' CCTGCACCTATAAACACAGAATAGA 3'), *ppm1j* (5' AATTTGTGACATCAGCGGCACGGTA 3'), *acp1* ATG (5' TCCGCTGGAAGCCGVVATATTGGTC 3'), *acp1* (exon 1) (5' TATAGCATTTCTTACCCAAGCACAC 3'), *speg* ATG (5' TCTTCTCTTCAGTAACTTTTCTCAT 3'), *edem3* (exon 1) (5' AGTCCTCACACAGACACATACCTCA 3'), and *eya4* ATG (5' CAGATCCTGTGTATTCTCCATCAGT). Zebrafish embryos were injected with various concentrations of MO (*sos2* ATG– 150 uM, *sos2* (exon 2) – 400 uM, *ppm1j* ATG, *acp1* ATG, and *acp1* (exon1) – 400uM, *edem3* ATG – 300uM, *eya4* ATG – 400 uM, *speg* ATG – 100uM) at the one-cell stage. We fixed embryos in 4% PFA at relevant developmental stages for analysis by *in situ* hybridization (http://zfin.org/ZFIN/Methods/ThisseProtocol.html). Distinct pronephros (embryonic kidney) structures were visualized using a series of established

markers: *pax2a* (global), *wt1a* (podocyte), *slc20a1a* (proximal tubule), and *slc13a3* (distal tubule). *acp1* probe was generated from zebrafish cDNA using the following primers: 5' TGGAGAATAGACAGTGCCGC 3' (forward 1) and 5' TTTTCACGCTGCTTGCCTTC 3' (reverse 1), and 5' GTGGAGAATAGACAGTGCCG 3' (forward 2) and 5' CAGGAAGGCTTTGCATC 3' (reverse 2) . *sos2* probe was generated from zebrafish cDNA using the following primers: 5' GTGTTCGAGGAAGGAGCACA 3' (forward) and 5' TGATGTTCCACCCACTGACG 3' (reverse).

Abnormal gene expression patterns were identified by direct comparison to control embryos that were injected with a standard control MO designed by GeneTools (SynGene, Cambridge, UK). Developmental phenotypes were scored by two independent researchers. Fisher's exact tests were used to test for normal and abnormal embryonic phenotypes for the *pax2a*, *wt1a* and *slc13a3* markers and Student-t test was used to test for differences in proximal tubule length for the *slc20a1a* marker; $P<0.05$ was set as a threshold for statistical significance. To evaluate proximal tubule length, the distance between the most anterior and the most posterior tip of the right proximal tubule (from standardized dorsal-view images) was measured using the imageJ measurement tool. Fisher's Exact test was used to test for differences in edema prevalence; $P<0.05$ was set as a threshold for statistical significance.

Dextran clearance experiments were performed following previously described protocol[12]. 72 hours after morpholino injection, embryos were anesthetized in a 1:20 dilution of 4mg/ml Tricane in embryo water and placed on

457 a 2% agarose injection mold. An equal volume of tetramethylrhodamine dextran
458 (70,000 MW; Invitrogen) was injected into the cardiac sinus venosus of each
459 embryo, and individual embryos were sorted into designated wells for timelapse
460 imaging. Fluorescent microscopy images were taken at 2 hpi (74 hpf), 24 hpi (96
461 hpf), and 48 hpi (120 hpf) for each sorted embryo to assess loading fluorescence
462 and the dextran clearance over time. Fluorescence intensity in the cardiac region
463 was measured as the mean grayscale value using Image J as previously
464 described[34]. Remaining fluorescence intensity at each time point was normalized
465 to the starting intensity per and plotted as a % of the initial fluorescence intensity.
466 To evaluate edema, morpholinos were injected into the single-cell stage embryo
467 and embryos were examined every 24 hours for evidence of edema. The number
468 of affected embryos was recorded as a fraction of total number of injected
469 embryos.
470
471 *Power calculation*
472 Power for association was evaluated for eGFRcrea assuming a mean of
473 4.5 ln(ml/min/1.73 m^2) with standard deviation of 0.2 ln(ml/min/1.73 m^2),
474 estimates from the European ancestry samples in the ARIC study, using
475 QUANTO power calculator, version 1.2.4 (http://biostats.usc.edu/Quanto.html) at
476 the significance level of 3.7×10^{-7} for a variant with MAF of 0.1, 0.05, 0.03, 0.01,
477 0.005, 0.001, or 0.0005. In Stage 1 with sample size n= 111,666, there was at
478 least 80% power to detect effect sizes of 0.008, 0.011, 0.015, 0.026, 0.036, 0.08,
479 or 0.11 ln(ml/min/1.73m^2), respectively. For Stage 2, using one-sided tests, there

480 was 92% power to detect an effect size of 0.018 for a variant with MAF of 0.02
481 (minimum MAF among all 8 SNPs tested) based on 48,343 samples in Stage 2
482 and correcting for the 8 SNPs tested.

Acknowledgements: The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services. Acknowledgements for all contributing studies are presented in the Supplementary Materials.

Statement of competing financial interests: None.

References:

1. Satko, SG, Sedor, JR, Iyengar, SK, Freedman, BI: Familial clustering of chronic kidney disease. *Seminars in dialysis,* 20**:** 229-236, 2007.
2. Boger, CA, Chen, MH, Tin, A, Olden, M, Kottgen, A, de Boer, IH, Fuchsberger, C, O'Seaghdha, CM, Pattaro, C, Teumer, A, Liu, CT, Glazer, NL, Li, M, O'Connell, JR, Tanaka, T, Peralta, CA, Kutalik, Z, Luan, J, Zhao, JH, Hwang, SJ, Akylbekova, E, Kramer, H, van der Harst, P, Smith, AV, Lohman, K, de Andrade, M, Hayward, C, Kollerits, B, Tonjes, A, Aspelund, T, Ingelsson, E, Eiriksdottir, G, Launer, LJ, Harris, TB, Shuldiner, AR, Mitchell, BD, Arking, DE, Franceschini, N, Boerwinkle, E, Egan, J, Hernandez, D, Reilly, M, Townsend, RR, Lumley, T, Siscovick, DS, Psaty, BM, Kestenbaum, B, Haritunians, T, Bergmann, S, Vollenweider, P, Waeber, G, Mooser, V, Waterworth, D, Johnson, AD, Florez, JC, Meigs, JB, Lu, X, Turner, ST, Atkinson, EJ, Leak, TS, Aasarod, K, Skorpen, F, Syvanen, AC, Illig, T, Baumert, J, Koenig, W, Kramer, BK, Devuyst, O, Mychaleckyj, JC, Minelli, C, Bakker, SJ, Kedenko, L, Paulweber, B, Coassin, S, Endlich, K, Kroemer, HK, Biffar, R, Stracke, S, Volzke, H, Stumvoll, M, Magi, R, Campbell, H, Vitart, V, Hastie, ND, Gudnason, V, Kardia, SL, Liu, Y, Polasek, O, Curhan, G, Kronenberg, F, Prokopenko, I, Rudan, I, Arnlov, J, Hallan, S, Navis, G, Consortium, CK, Parsa, A, Ferrucci, L, Coresh, J, Shlipak, MG, Bull, SB, Paterson, NJ, Wichmann, HE, Wareham, NJ, Loos, RJ, Rotter, JI, Pramstaller, PP, Cupples, LA, Beckmann, JS, Yang, Q, Heid, IM, Rettig, R, Dreisbach, AW, Bochud, M, Fox, CS, Kao, WH: CUBN is a gene locus for albuminuria. *Journal of the American Society of Nephrology : JASN,* 22**:** 555-570, 2011.
3. Kottgen, A, Pattaro, C, Boger, CA, Fuchsberger, C, Olden, M, Glazer, NL, Parsa, A, Gao, X, Yang, Q, Smith, AV, O'Connell, JR, Li, M, Schmidt, H, Tanaka, T, Isaacs, A, Ketkar, S, Hwang, SJ, Johnson, AD, Dehghan, A, Teumer, A, Pare, G, Atkinson, EJ, Zeller, T, Lohman, K, Cornelis, MC, Probst-Hensch, NM, Kronenberg, F, Tonjes, A, Hayward, C, Aspelund, T, Eiriksdottir, G, Launer, LJ, Harris, TB, Rampersaud, E, Mitchell, BD, Arking, DE, Boerwinkle, E, Struchalin, M, Cavalieri, M, Singleton, A, Giallauria, F, Metter, J, de Boer, IH, Haritunians, T, Lumley, T, Siscovick, D, Psaty, BM, Zillikens, MC, Oostra, BA, Feitosa, M, Province, M, de Andrade, M, Turner, ST, Schillert, A, Ziegler, A, Wild, PS, Schnabel, RB, Wilde, S, Munzel, TF, Leak, TS, Illig, T, Klopp, N, Meisinger, C, Wichmann, HE, Koenig, W, Zgaga, L, Zemunik, T, Kolcic, I, Minelli, C, Hu, FB, Johansson, A, Igl, W, Zaboli, G, Wild, SH, Wright, AF, Campbell, H, Ellinghaus, D, Schreiber, S, Aulchenko, YS, Felix, JF, Rivadeneira, F, Uitterlinden, AG, Hofman, A, Imboden, M, Nitsch, D, Brandstatter, A, Kollerits, B, Kedenko, L, Magi, R, Stumvoll, M, Kovacs, P, Boban, M, Campbell, S, Endlich, K, Volzke, H, Kroemer, HK, Nauck, M, Volker, U, Polasek, O, Vitart, V, Badola, S, Parker, AN, Ridker, PM, Kardia, SL, Blankenberg, S, Liu, Y, Curhan, GC, Franke, A, Rochat, T, Paulweber, B, Prokopenko, I, Wang, W, Gudnason, V, Shuldiner, AR, Coresh, J,

Schmidt, R, Ferrucci, L, Shlipak, MG, van Duijn, CM, Borecki, I, Kramer, BK, Rudan, I, Gyllensten, U, Wilson, JF, Witteman, JC, Pramstaller, PP, Rettig, R, Hastie, N, Chasman, DI, Kao, WH, Heid, IM, Fox, CS: New loci associated with kidney function and chronic kidney disease. *Nature genetics,* 42**:** 376-384, 2010.

4. Pattaro, C, Kottgen, A, Teumer, A, Garnaas, M, Boger, CA, Fuchsberger, C, Olden, M, Chen, MH, Tin, A, Taliun, D, Li, M, Gao, X, Gorski, M, Yang, Q, Hundertmark, C, Foster, MC, O'Seaghdha, CM, Glazer, N, Isaacs, A, Liu, CT, Smith, AV, O'Connell, JR, Struchalin, M, Tanaka, T, Li, G, Johnson, AD, Gierman, HJ, Feitosa, M, Hwang, SJ, Atkinson, EJ, Lohman, K, Cornelis, MC, Johansson, A, Tonjes, A, Dehghan, A, Chouraki, V, Holliday, EG, Sorice, R, Kutalik, Z, Lehtimaki, T, Esko, T, Deshmukh, H, Ulivi, S, Chu, AY, Murgia, F, Trompet, S, Imboden, M, Kollerits, B, Pistis, G, Consortium, CA, Consortium, I, Consortium, CA, Wellcome Trust Case Control, C, Harris, TB, Launer, LJ, Aspelund, T, Eiriksdottir, G, Mitchell, BD, Boerwinkle, E, Schmidt, H, Cavalieri, M, Rao, M, Hu, FB, Demirkan, A, Oostra, BA, de Andrade, M, Turner, ST, Ding, J, Andrews, JS, Freedman, BI, Koenig, W, Illig, T, Doring, A, Wichmann, HE, Kolcic, I, Zemunik, T, Boban, M, Minelli, C, Wheeler, HE, Igl, W, Zaboli, G, Wild, SH, Wright, AF, Campbell, H, Ellinghaus, D, Nothlings, U, Jacobs, G, Biffar, R, Endlich, K, Ernst, F, Homuth, G, Kroemer, HK, Nauck, M, Stracke, S, Volker, U, Volzke, H, Kovacs, P, Stumvoll, M, Magi, R, Hofman, A, Uitterlinden, AG, Rivadeneira, F, Aulchenko, YS, Polasek, O, Hastie, N, Vitart, V, Helmer, C, Wang, JJ, Ruggiero, D, Bergmann, S, Kahonen, M, Viikari, J, Nikopensius, T, Province, M, Ketkar, S, Colhoun, H, Doney, A, Robino, A, Giulianini, F, Kramer, BK, Portas, L, Ford, I, Buckley, BM, Adam, M, Thun, GA, Paulweber, B, Haun, M, Sala, C, Metzger, M, Mitchell, P, Ciullo, M, Kim, SK, Vollenweider, P, Raitakari, O, Metspalu, A, Palmer, C, Gasparini, P, Pirastu, M, Jukema, JW, Probst-Hensch, NM, Kronenberg, F, Toniolo, D, Gudnason, V, Shuldiner, AR, Coresh, J, Schmidt, R, Ferrucci, L, Siscovick, DS, van Duijn, CM, Borecki, I, Kardia, SL, Liu, Y, Curhan, GC, Rudan, I, Gyllensten, U, Wilson, JF, Franke, A, Pramstaller, PP, Rettig, R, Prokopenko, I, Witteman, JC, Hayward, C, Ridker, P, Parsa, A, Bochud, M, Heid, IM, Goessling, W, Chasman, DI, Kao, WH, Fox, CS: Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS genetics,* 8**:** e1002584, 2012.

5. Okada, Y, Sim, X, Go, MJ, Wu, JY, Gu, D, Takeuchi, F, Takahashi, A, Maeda, S, Tsunoda, T, Chen, P, Lim, SC, Wong, TY, Liu, J, Young, TL, Aung, T, Seielstad, M, Teo, YY, Kim, YJ, Lee, JY, Han, BG, Kang, D, Chen, CH, Tsai, FJ, Chang, LC, Fann, SJ, Mei, H, Rao, DC, Hixson, JE, Chen, S, Katsuya, T, Isono, M, Ogihara, T, Chambers, JC, Zhang, W, Kooner, JS, KidneyGen, C, Consortium, CK, Albrecht, E, consortium, G, Yamamoto, K, Kubo, M, Nakamura, Y, Kamatani, N, Kato, N, He, J, Chen, YT, Cho, YS, Tai, ES, Tanaka, T: Meta-analysis identifies multiple loci associated with

kidney function-related traits in east Asian populations. *Nature genetics,* 44**:** 904-909, 2012.

6. Pattaro, C, Teumer, A, Gorski, M, Chu, AY, Li, M, Mijatovic, V, Garnaas, M, Tin, A, Sorice, R, Li, Y, Taliun, D, Olden, M, Foster, M, Yang, Q, Chen, M-H, Pers, TH, Johnson, AD, Ko, Y-A, Fuchsberger, C, Tayo, B, Nalls, M, Feitosa, MF, Isaacs, A, Dehghan, A, d/'Adamo, P, Adeyemo, A, Dieffenbach, AK, Zonderman, AB, Nolte, IM, van der Most, PJ, Wright, AF, Shuldiner, AR, Morrison, AC, Hofman, A, Smith, AV, Dreisbach, AW, Franke, A, Uitterlinden, AG, Metspalu, A, Tonjes, A, Lupo, A, Robino, A, Johansson, A, Demirkan, A, Kollerits, B, Freedman, BI, Ponte, B, Oostra, BA, Paulweber, B, Kramer, BK, Mitchell, BD, Buckley, BM, Peralta, CA, Hayward, C, Helmer, C, Rotimi, CN, Shaffer, CM, Muller, C, Sala, C, van Duijn, CM, Saint-Pierre, A, Ackermann, D, Shriner, D, Ruggiero, D, Toniolo, D, Lu, Y, Cusi, D, Czamara, D, Ellinghaus, D, Siscovick, DS, Ruderfer, D, Gieger, C, Grallert, H, Rochtchina, E, Atkinson, EJ, Holliday, EG, Boerwinkle, E, Salvi, E, Bottinger, EP, Murgia, F, Rivadeneira, F, Ernst, F, Kronenberg, F, Hu, FB, Navis, GJ, Curhan, GC, Ehret, GB, Homuth, G, Coassin, S, Thun, G-A, Pistis, G, Gambaro, G, Malerba, G, Montgomery, GW, Eiriksdottir, G, Jacobs, G, Li, G, Wichmann, HE, Campbell, H, Schmidt, H, Wallaschofski, H, Volzke, H, Brenner, H, Kroemer, HK, Kramer, H, Lin, H, Leach, IM, Ford, I, Guessous, I, Rudan, I, Prokopenko, I, Borecki, I, Heid, IM, Kolcic, I, Persico, I, Jukema, JW, Wilson, JF, Felix, JF, Divers, J, Lambert, J-C, Stafford, JM, Gaspoz, J-M, Smith, JA, Faul, JD, Wang, JJ, Ding, J, Hirschhorn, JN, Attia, J, Whitfield, JB, Chalmers, J, Viikari, J, Coresh, J, Denny, JC, Karjalainen, J, Fernandes, JK, Endlich, K, Butterbach, K, Keene, KL, Lohman, K, Portas, L, Launer, LJ, Lyytikainen, L-P, Yengo, L, Franke, L, Ferrucci, L, Rose, LM, Kedenko, L, Rao, M, Struchalin, M, Kleber, ME, Cavalieri, M, Haun, M, Cornelis, MC, Ciullo, M, Pirastu, M, de Andrade, M, McEvoy, MA, Woodward, M, Adam, M, Cocca, M, Nauck, M, Imboden, M, Waldenberger, M, Pruijm, M, Metzger, M, Stumvoll, M, Evans, MK, Sale, MM, Kahonen, M, Boban, M, Bochud, M, Rheinberger, M, Verweij, N, Bouatia-Naji, N, Martin, NG, Hastie, N, Probst-Hensch, N, Soranzo, N, Devuyst, O, Raitakari, O, Gottesman, O, Franco, OH, Polasek, O, Gasparini, P, Munroe, PB, Ridker, PM, Mitchell, P, Muntner, P, Meisinger, C, Smit, JH, Consortium, I, Consortium, A, Cardiogram, Group, CH-HF, Consortium, EC, Kovacs, P, Wild, PS, Froguel, P, Rettig, R, Magi, R, Biffar, R, Schmidt, R, Middelberg, RPS, Carroll, RJ, Penninx, BW, Scott, RJ, Katz, R, Sedaghat, S, Wild, SH, Kardia, SLR, Ulivi, S, Hwang, S-J, Enroth, S, Kloiber, S, Trompet, S, Stengel, B, Hancock, SJ, Turner, ST, Rosas, SE, Stracke, S, Harris, TB, Zeller, T, Zemunik, T, Lehtimaki, T, Illig, T, Aspelund, T, Nikopensius, T, Esko, T, Tanaka, T, Gyllensten, U, Volker, U, Emilsson, V, Vitart, V, Aalto, V, Gudnason, V, Chouraki, V, Chen, W-M, Igl, W, Marz, W, Koenig, W, Lieb, W, Loos, RJF, Liu, Y, Snieder, H, Pramstaller, PP, Parsa, A, O/'Connell, JR, Susztak, K, Hamet, P, Tremblay, J, de Boer, IH, Boger, CA, Goessling, W, Chasman, DI,

Kottgen, A, Kao, WHL, Fox, CS: Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun,* 7, 2016.
7. Cornelis, MC, Monda, KL, Yu, K, Paynter, N, Azzato, EM, Bennett, SN, Berndt, SI, Boerwinkle, E, Chanock, S, Chatterjee, N, Couper, D, Curhan, G, Heiss, G, Hu, FB, Hunter, DJ, Jacobs, K, Jensen, MK, Kraft, P, Landi, MT, Nettleton, JA, Purdue, MP, Rajaraman, P, Rimm, EB, Rose, LM, Rothman, N, Silverman, D, Stolzenberg-Solomon, R, Subar, A, Yeager, M, Chasman, DI, van Dam, RM, Caporaso, NE: Genome-Wide Meta-Analysis Identifies Regions on 7p21 (<italic>AHR</italic>) and 15q24 (<italic>CYP1A2</italic>) As Determinants of Habitual Caffeine Consumption. *PLoS Genet,* 7**:** e1002033, 2011.
8. Sulem, P, Gudbjartsson, DF, Geller, F, Prokopenko, I, Feenstra, B, Aben, KKH, Franke, B, den Heijer, M, Kovacs, P, Stumvoll, M, Mägi, R, Yanek, LR, Becker, LC, Boyd, HA, Stacey, SN, Walters, GB, Jonasdottir, A, Thorleifsson, G, Holm, H, Gudjonsson, SA, Rafnar, T, Björnsdottir, G, Becker, DM, Melbye, M, Kong, A, Tönjes, A, Thorgeirsson, T, Thorsteinsdottir, U, Kiemeney, LA, Stefansson, K: Sequence variants at CYP1A1–CYP1A2 and AHR associate with coffee consumption. *Human Molecular Genetics,* 20**:** 2071-2077, 2011.
9. Imielinski, M, Baldassano, RN, Griffiths, A, Russell, RK, Annese, V, Dubinsky, M, Kugathasan, S, Bradfield, JP, Walters, TD, Sleiman, P, Kim, CE, Muise, A, Wang, K, Glessner, JT, Saeed, S, Zhang, H, Frackelton, EC, Hou, C, Flory, JH, Otieno, G, Chiavacci, RM, Grundmeier, R, Castro, M, Latiano, A, Dallapiccola, B, Stempak, J, Abrams, DJ, Taylor, K, McGovern, D, Silber, G, Wrobel, I, Quiros, A, Barrett, JC, Hansoul, S, Nicolae, DL, Cho, JH, Duerr, RH, Rioux, JD, Brant, SR, Silverberg, MS, Taylor, KD, Barmuda, MM, Bitton, A, Dassopoulos, T, Datta, LW, Green, T, Griffiths, AM, Kistner, EO, Murtha, MT, Regueiro, MD, Rotter, JI, Schumm, LP, Steinhart, AH, Targan, SR, Xavier, RJ, Libioulle, C, Sandor, C, Lathrop, M, Belaiche, J, Dewit, O, Gut, I, Heath, S, Laukens, D, Mni, M, Rutgeerts, P, Van Gossum, A, Zelenika, D, Franchimont, D, Hugot, JP, de Vos, M, Vermeire, S, Louis, E, Cardon, LR, Anderson, CA, Drummond, H, Nimmo, E, Ahmad, T, Prescott, NJ, Onnie, CM, Fisher, SA, Marchini, J, Ghori, J, Bumpstead, S, Gwillam, R, Tremelling, M, Delukas, P, Mansfield, J, Jewell, D, Satsangi, J, Mathew, CG, Parkes, M, Georges, M, Daly, MJ, Heyman, MB, Ferry, GD, Kirschner, B, Lee, J, Essers, J, Grand, R, Stephens, M, Levine, A, Piccoli, D, Van Limbergen, J, Cucchiara, S, Monos, DS, Guthery, SL, Denson, L, Wilson, DC, Grant, SF, Daly, M, Silverberg, MS, Satsangi, J, Hakonarson, H: Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet,* 41**:** 1335-1340, 2009.
10. Uhlen, M, Fagerberg, L, Hallstrom, BM, Lindskog, C, Oksvold, P, Mardinoglu, A, Sivertsson, A, Kampf, C, Sjostedt, E, Asplund, A, Olsson, I, Edlund, K, Lundberg, E, Navani, S, Szigyarto, CA, Odeberg, J, Djureinovic, D, Takanen, JO, Hober, S, Alm, T, Edqvist, PH, Berling, H, Tegel, H, Mulder,

J, Rockberg, J, Nilsson, P, Schwenk, JM, Hamsten, M, von Feilitzen, K, Forsberg, M, Persson, L, Johansson, F, Zwahlen, M, von Heijne, G, Nielsen, J, Ponten, F: Proteomics. Tissue-based map of the human proteome. *Science,* 347**:** 1260419, 2015.

11. Hanke, N, Staggs, L, Schroder, P, Litteral, J, Fleig, S, Kaufeld, J, Pauli, C, Haller, H, Schiffer, M: "Zebrafishing" for novel genes relevant to the glomerular filtration barrier. *BioMed research international,* 2013**:** 658270, 2013.
12. Hentschel, DM, Mengel, M, Boehme, L, Liebsch, F, Albertin, C, Bonventre, JV, Haller, H, Schiffer, M: Rapid screening of glomerular slit diaphragm integrity in larval zebrafish. *American journal of physiology Renal physiology,* 293**:** F1746-1750, 2007.
13. Liu, C, Kraja, A, Smith, JA, Brody, JA, Franceschini, N, Bis, JC, Rice, K, Morrison, AC, Lu, Y, Weiss, S, Guo, X, Palmas, W, Martin, LW, Chen, YI, Surendran, P, Drenos, F, Cook, JP, Ehret, G, Newton-Cheh, C, Levy, D, Chasman, DI: An exome-centered association analysis identifies novel common and rare blood pressure variants with enrichment for genes involved in cardiometabolic phenotypes. *Manuscript resubmitted for publication at Nature Genetics,* Reference to be updated upon publication, 2016.
14. Wessel, J, Chu, AY, Willems, SM, Wang, S, Yaghootkar, H, Brody, JA, Dauriz, M, Hivert, M-F, Raghavan, S, Lipovich, L, Hidalgo, B, Fox, K, Huffman, JE, An, P, Lu, Y, Rasmussen-Torvik, LJ, Grarup, N, Ehm, MG, Li, L, Baldridge, AS, Stancakova, A, Abrol, R, Besse, C, Boland, A, Bork-Jensen, J, Fornage, M, Freitag, DF, Garcia, ME, Guo, X, Hara, K, Isaacs, A, Jakobsdottir, J, Lange, LA, Layton, JC, Li, M, Hua Zhao, J, Meidtner, K, Morrison, AC, Nalls, MA, Peters, MJ, Sabater-Lleal, M, Schurmann, C, Silveira, A, Smith, AV, Southam, L, Stoiber, MH, Strawbridge, RJ, Taylor, KD, Varga, TV, Allin, KH, Amin, N, Aponte, JL, Aung, T, Barbieri, C, Bihlmeyer, NA, Boehnke, M, Bombieri, C, Bowden, DW, Burns, SM, Chen, Y, Chen, Y-D, Cheng, C-Y, Correa, A, Czajkowski, J, Dehghan, A, Ehret, GB, Eiriksdottir, G, Escher, SA, Farmaki, A-E, Franberg, M, Gambaro, G, Giulianini, F, Goddard, WA, Goel, A, Gottesman, O, Grove, ML, Gustafsson, S, Hai, Y, Hallmans, G, Heo, J, Hoffmann, P, Ikram, MK, Jensen, RA, Jorgensen, ME, Jorgensen, T, Karaleftheri, M, Khor, CC, Kirkpatrick, A, Kraja, AT, Kuusisto, J, Lange, EM, Lee, IT, Lee, W-J, Leong, A, Liao, J, Liu, C, Liu, Y, Lindgren, CM, Linneberg, A, Malerba, G, Mamakou, V, Marouli, E, Maruthur, NM, Matchan, A, McKean-Cowdin, R, McLeod, O, Metcalf, GA, Mohlke, KL, Muzny, DM, Ntalla, I, Palmer, ND, Pasko, D, Peter, A, Rayner, NW, Renstrom, F, Rice, K, Sala, CF, Sennblad, B, Serafetinidis, I, Smith, JA, Soranzo, N, Speliotes, EK, Stahl, EA, Stirrups, K, Tentolouris, N, Thanopoulou, A, Torres, M, Traglia, M, Tsafantakis, E, Javad, S, Yanek, LR, Zengini, E, Becker, DM, Bis, JC, Brown, JB, Adrienne Cupples, L, Hansen, T, Ingelsson, E, Karter, AJ, Lorenzo, C, Mathias, RA, Norris, JM, Peloso, GM, Sheu, WHH, Toniolo, D, Vaidya, D, Varma, R, Wagenknecht, LE, Boeing, H, Bottinger, EP,

719 Dedoussis, G, Deloukas, P, Ferrannini, E, Franco, OH, Franks, PW,
720 Gibbs, RA, Gudnason, V, Hamsten, A, Harris, TB, Hattersley, AT,
721 Hayward, C, Hofman, A, Jansson, J-H, Langenberg, C, Launer, LJ, Levy,
722 D, Oostra, BA, O/'Donnell, CJ, O/'Rahilly, S, Padmanabhan, S, Pankow,
723 JS, Polasek, O, Province, MA, Rich, SS, Ridker, PM, Rudan, I, Schulze,
724 MB, Smith, BH, Uitterlinden, AG, Walker, M, Watkins, H, Wong, TY,
725 Zeggini, E, The, E-IC, Laakso, M, Borecki, IB, Chasman, DI, Pedersen, O,
726 Psaty, BM, Shyong Tai, E, van Duijn, CM, Wareham, NJ, Waterworth, DM,
727 Boerwinkle, E, Linda Kao, WH, Florez, JC, Loos, RJF, Wilson, JG,
728 Frayling, TM, Siscovick, DS, Dupuis, J, Rotter, JI, Meigs, JB, Scott, RA,
729 Goodarzi, MO: Low-frequency and rare exome chip variants associate
730 with fasting glucose and type 2 diabetes susceptibility. *Nat Commun,* 6,
731 2015.
732 15. Pattaro, C, Teumer, A, Gorski, M, Chu, AY, Li, M, Mijatovic, V, Garnaas, M,
733 Tin, A, Sorice, R, Li, Y, Taliun, D, Olden, M, Foster, MC, Yang, Q, Chen,
734 M, Pers, TH, Johnson, AD, Ko, Y, Fuchsberger, C, Tayo, B, Nalls, M,
735 Feitosa, M, Boger, CA, Goessling, W, Chasman, DI, Kottgen, A, Kao, WH,
736 Fox, CS: Genetic associations at 53 loci highlight cell types and biological
737 pathways relevant for kidney function. *Manuscript in press at Nature*
738 *Communications,* Confidential draft of manuscript included in reference
739 materials., 2016.
740 16. Yamamoto, GL, Aguena, M, Gos, M, Hung, C, Pilch, J, Fahiminiya, S,
741 Abramowicz, A, Cristian, I, Buscarilli, M, Naslavsky, MS, Malaquias, AC,
742 Zatz, M, Bodamer, O, Majewski, J, Jorge, AAL, Pereira, AC, Kim, CA,
743 Passos-Bueno, MR, Bertola, DR: Rare variants in SOS2 and LZTR1 are
744 associated with Noonan syndrome. *Journal of Medical Genetics,* 52**:** 413-
745 421, 2015.
746 17. Romano, AA, Allanson, JE, Dahlgren, J, Gelb, BD, Hall, B, Pierpont, ME,
747 Roberts, AE, Robinson, W, Takemoto, CM, Noonan, JA: Noonan
748 syndrome: clinical features, diagnosis, and management guidelines.
749 *Pediatrics,* 126**:** 746-759, 2010.
750 18. George, CD, Patton, MA, el Sawi, M, Sharland, M, Adam, EJ: Abdominal
751 ultrasound in Noonan syndrome: a study of 44 patients. *Pediatric*
752 *radiology,* 23**:** 316-318, 1993.
753 19. Mancini, F, Rigacci, S, Berti, A, Balduini, C, Torti, M: The low-molecular-
754 weight phosphotyrosine phosphatase is a negative regulator of
755 FcgammaRIIA-mediated cell activation. *Blood,* 110**:** 1871-1878, 2007.
756 20. Peloso, GM, Auer, PL, Bis, JC, Voorman, A, Morrison, AC, Stitziel, NO,
757 Brody, JA, Khetarpal, SA, Crosby, JR, Fornage, M, Isaacs, A,
758 Jakobsdottir, J, Feitosa, MF, Davies, G, Huffman, JE, Manichaikul, A,
759 Davis, B, Lohman, K, Joon, AY, Smith, AV, Grove, ML, Zanoni, P, Redon,
760 V, Demissie, S, Lawson, K, Peters, U, Carlson, C, Jackson, RD,
761 Ryckman, KK, Mackey, RH, Robinson, JG, Siscovick, DS, Schreiner, PJ,
762 Mychaleckyj, JC, Pankow, JS, Hofman, A, Uitterlinden, AG, Harris, TB,
763 Taylor, KD, Stafford, JM, Reynolds, LM, Marioni, RE, Dehghan, A, Franco,
764 OH, Patel, AP, Lu, Y, Hindy, G, Gottesman, O, Bottinger, EP, Melander,

O, Orho-Melander, M, Loos, RJ, Duga, S, Merlini, PA, Farrall, M, Goel, A, Asselta, R, Girelli, D, Martinelli, N, Shah, SH, Kraus, WE, Li, M, Rader, DJ, Reilly, MP, McPherson, R, Watkins, H, Ardissino, D, Zhang, Q, Wang, J, Tsai, MY, Taylor, HA, Correa, A, Griswold, ME, Lange, LA, Starr, JM, Rudan, I, Eiriksdottir, G, Launer, LJ, Ordovas, JM, Levy, D, Chen, YD, Reiner, AP, Hayward, C, Polasek, O, Deary, IJ, Borecki, IB, Liu, Y, Gudnason, V, Wilson, JG, van Duijn, CM, Kooperberg, C, Rich, SS, Psaty, BM, Rotter, JI, O'Donnell, CJ, Rice, K, Boerwinkle, E, Kathiresan, S, Cupples, LA: Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet,* 94**:** 223-232, 2014.

21. Auer, PL, Teumer, A, Schick, U, O'Shaughnessy, A, Lo, KS, Chami, N, Carlson, C, de Denus, S, Dube, MP, Haessler, J, Jackson, RD, Kooperberg, C, Perreault, LP, Nauck, M, Peters, U, Rioux, JD, Schmidt, F, Turcot, V, Volker, U, Volzke, H, Greinacher, A, Hsu, L, Tardif, JC, Diaz, GA, Reiner, AP, Lettre, G: Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nat Genet,* 46**:** 629-634, 2014.
22. Huffman, JE, de Vries, PS, Morrison, AC, Sabater-Lleal, M, Kacprowski, T, Auer, PL, Brody, JA, Chasman, DI, Chen, MH, Guo, X, Lin, LA, Marioni, RE, Muller-Nurasyid, M, Yanek, LR, Pankratz, N, Grove, ML, de Maat, MP, Cushman, M, Wiggins, KL, Qi, L, Sennblad, B, Harris, SE, Polasek, O, Riess, H, Rivadeneira, F, Rose, LM, Goel, A, Taylor, KD, Teumer, A, Uitterlinden, AG, Vaidya, D, Yao, J, Tang, W, Levy, D, Waldenberger, M, Becker, DM, Folsom, AR, Giulianini, F, Greinacher, A, Hofman, A, Huang, CC, Kooperberg, C, Silveira, A, Starr, JM, Strauch, K, Strawbridge, RJ, Wright, AF, McKnight, B, Franco, OH, Zakai, N, Mathias, RA, Psaty, BM, Ridker, PM, Tofler, GH, Volker, U, Watkins, H, Fornage, M, Hamsten, A, Deary, IJ, Boerwinkle, E, Koenig, W, Rotter, JI, Hayward, C, Dehghan, A, Reiner, AP, O'Donnell, CJ, Smith, NL: Rare and low-frequency variants and their association with plasma levels of fibrinogen, FVII, FVIII, and vWF. *Blood,* 126**:** e19-29, 2015.
23. Fox, CS, Larson, MG, Leip, EP, Culleton, B, Wilson, PW, Levy, D: Predictors of new-onset kidney disease in a community-based population. *JAMA : the journal of the American Medical Association,* 291**:** 844-850, 2004.
24. Selvin, E, Manzi, J, Stevens, LA, Van Lente, F, Lacher, DA, Levey, AS, Coresh, J: Calibration of serum creatinine in the National Health and Nutrition Examination Surveys (NHANES) 1988-1994, 1999-2004. *American Journal of Kidney Diseases : The Official Journal of the National Kidney Foundation,* 50**:** 918-926, 2007.
25. Levey, AS, de Jong, PE, Coresh, J, El Nahas, M, Astor, BC, Matsushita, K, Gansevoort, RT, Kasiske, BL, Eckardt, KU: The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney international,* 80**:** 17-28, 2011.
26. Goldstein, JI, Crenshaw, A, Carey, J, Grant, GB, Maguire, J, Fromer, M, O'Dushlaine, C, Moran, JL, Chambert, K, Stevens, C, Swedish

Schizophrenia, C, Consortium, AAS, Sklar, P, Hultman, CM, Purcell, S, McCarroll, SA, Sullivan, PF, Daly, MJ, Neale, BM: zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics (Oxford, England),* 28**:** 2543-2545, 2012.
27. Grove, ML, Yu, B, Cochran, BJ, Haritunians, T, Bis, JC, Taylor, KD, Hansen, M, Borecki, IB, Cupples, LA, Fornage, M, Gudnason, V, Harris, TB, Kathiresan, S, Kraaij, R, Launer, LJ, Levy, D, Liu, Y, Mosley, T, Peloso, GM, Psaty, BM, Rich, SS, Rivadeneira, F, Siscovick, DS, Smith, AV, Uitterlinden, A, van Duijn, CM, Wilson, JG, O'Donnell, CJ, Rotter, JI, Boerwinkle, E: Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PloS one,* 8**:** e68095, 2013.
28. Team, RC: R: A Language and Environment for Statistical Computing. 2008.
29. Willer, CJ, Li, Y, Abecasis, GR: METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England),* 26**:** 2190-2191, 2010.
30. Li, B, Leal, SM: Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American Journal of Human Genetics,* 83**:** 311-321, 2008.
31. Voorman, A, Brody, J, Chen, H, Lumley, T: seqMeta: An R package for meta-analyzing region-based tests of rare DNA variants. . *R Package version 14*, 2014.
32. Wu, MC, Lee, S, Cai, T, Li, Y, Boehnke, M, Lin, X: Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics,* 89**:** 82-93, 2011.
33. Liu, X, Jian, X, Boerwinkle, E: dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Human mutation,* 34**:** E2393-2402, 2013.
34. Christou-Savina, S, Beales, PL, Osborn, DP: Evaluation of zebrafish kidney function using a fluorescent clearance assay. *Journal of visualized experiments : JoVE***:** e52540, 2015.

Figure Legends

Figure 1. Manhattan plot for single-variant analysis in eGFRcrea among 111666 European ancestry participants. Newly identified variants are in dark orange. The gene, *ORC4*, not successfully replicated, are in orange. Known loci are in blue.

Figure 2. ***sos2* and *acp1* knockdowns result in defective kidney development. A-D).** Whole mount *in situ* hybridization in control embryo demonstrates normal expression of kidney markers, including *pax2a* (global kidney, A), *slc20a1a* (proximal tubules, B), and *slc12a3* (distal tubules, C) at 48 hours post fertilization (hpf), and *wt1a* (podocytes, D) at 24 hpf. **E-L)** *sos2* and *acp1* ATG morpholino (MO) knockdown embryos develop glomerular gene expression defects (E, I, arrowheads) and display elongated proximal tubules (F, J). Knockdown of *acp1* shortened the distal tubules, while *sos2* knockdown left distal tubule *slc12a3* expression unaffected (G, K). No abnormalities in podocyte marker *wt1a* were observed for *sos2* ATG- and *acp1* ATG-MOs (H, L). **M)** Quantitative assessment of proximal tubule length (*slc20a1a* expression) shows that proximal tubules are elongated in *sos2* ATG- and *acp1* ATG-MO injected embryos. Student's t-test used to calculate p-values (P). **N)** Table of observed abnormal embryos and total number examined by kidney markers *pax2a*, *wt1a* and *slc12a3* and MO-injected or control status. Fisher's exact test used to calculate p-values (P).

Figure 3. *sos2* and *acp1* knockdowns result in altered kidney function. A) *sos2* and *acp1* morphants develop edema, which is a sign of kidney failure in zebrafish. *sos2* morphants display severe global edema at 120 hpf, with fluid accumulation in the pericardium (black arrow) and intestinal tract (black star). At 96 hpf, *acp1* morphants have severe pericardial edema (black arrow). **B-C)** Incidence of edema and embryonic lethality in *acp1* and *sos2* morphants with Fisher's Exact test. * indicates p-value <0.05; ** indicates p-value <0.005; *** indicates p-value <0.0001. **D-I)** Embryos were injected with control, *sos2*, or *acp1* morpholino at the single cell stage and subsequently injected with 70,000 MW fluorescent rhodamine dextran at 72 hpf. Dextran fluorescence intensity was measured over 48 hours post injection. Dextran-injected embryos show equal loading at 2 hpi. Compared to control embryos, *sos2* and *acp1* MO injected embryos have reduced dextran clearance in the cardiovascular region over time. **J-L)** Morphant embryos have altered convoluted tubule morphology at 120 hpi (48 hpi) (white arrows). **M)** Dextran fluorescence intensity over time as normalized to starting fluorescence intensity.

Table 1. Novel variants associated with eGFRcrea in European ancestry participants from single-variant analysis in Stage 1 meeting chip-wide significance ($P<3.7x10^{-7}$) and associations in Stage 2 and combined analysis.

					Stage 1[c]				Stage 2[d]		Combined		
Locus[a]	**dbSNPID**	**Chr**	**Position[b]**	**Variation (Substitution)**	**EA/ Non-EA (EAF)**	**Beta (SE)**	**P**	**I^2**	**Beta (SE)**	**1-sided P**	**Beta (SE)**	**P**	**Prop Var Exp (%)**
PPM1J	rs34611728	1	113255456	c.639G>T (L213F)	A/C (0.13)	-0.0103 (0.0013)	1.2E-14	13.2	-0.0059 (0.0023)	4.7E-03	-0.0092 (0.0011)	3.3E-16	0.05
EDEM3	rs78444298	1	184672098	c.2236C>T (P746S)	A/G (0.02)	-0.0183 (0.0034)	5.2E-08	15.3	-0.0225 (0.0055)	1.8E-05	-0.0195 (0.0029)	1.5E-11	0.03
ACP1	rs11553746	2	272203	c.129C>T (T95I)	T/C (0.35)	-0.0049 (0.0009)	2.0E-07	20.7	-0.0032 (0.0016)	2.2E-02	-0.0045 (0.0008)	1.0E-08	0.02
ORC4[e]	rs2307394	2	148716428	c.233A>G (N78S)	C/T (0.32)	-0.0058 (0.0010)	6.8E-09	14.3	-0.0025 (0.0016)	6.0E-02	-0.0049 (0.0009)	8.4E-09	0.03
SPEG	rs55760516	2	220354108	c.8191A>G (R2731G)	G/A (0.33)	0.0059 (0.0009)	4.8E-10	0.5	0.0054 (0.0016)	3.7E-04	0.0058 (0.0008)	1.7E-13	0.04
EYA4	rs9493627	6	133789728	c.829G>A (G223S)	A/G (0.31)	0.0061 (0.0010)	2.3E-10	0.0	0.0049 (0.0016)	1.4E-03	0.0058 (0.0009)	1.4E-11	0.04
CYP1A1	rs2472297	15	75027880	intergenic	T/C (0.24)	0.0057 (0.0010)	7.0E-08	0.0	0.0059 (0.0017)	3.2E-04	0.0058 (0.0009)	3.0E-11	0.03
ATXN2L	rs8049439	16	28837515	intronic	C/T (0.40)	0.0048 (0.0009)	1.3E-07	7.1	0.0045 (0.0016)	1.8E-03	0.0047 (0.0008)	1.2E-09	0.03

[a] Loci are named according to the closest gene based on the position of the lead SNP for new loci.
[b] Position is reported in UCSC Genome Browser build hg19.
[c] Sample size for Stage 1 analysis: N=111666.
[d] Sample size for Stage 2 analysis: N=48343.
[e] This variant reached chip-wide significance ($P<3.7x10^{-7}$) in the Stage 1 samples but did not meet validation criteria in Stage 2.
Abbreviations: EA = effect allele, EAF = effect allele frequency, Chr = chromosome, SE = standard error, P= p-value, PropVarExp=proportion of variance in ln(eGFRcrea) explained.

Table 2. Genes associated with eGFRcrea in European ancestry participants from gene-based analyses meeting chip-wide significance thresholds ($P<2.5x10^{-6}$).

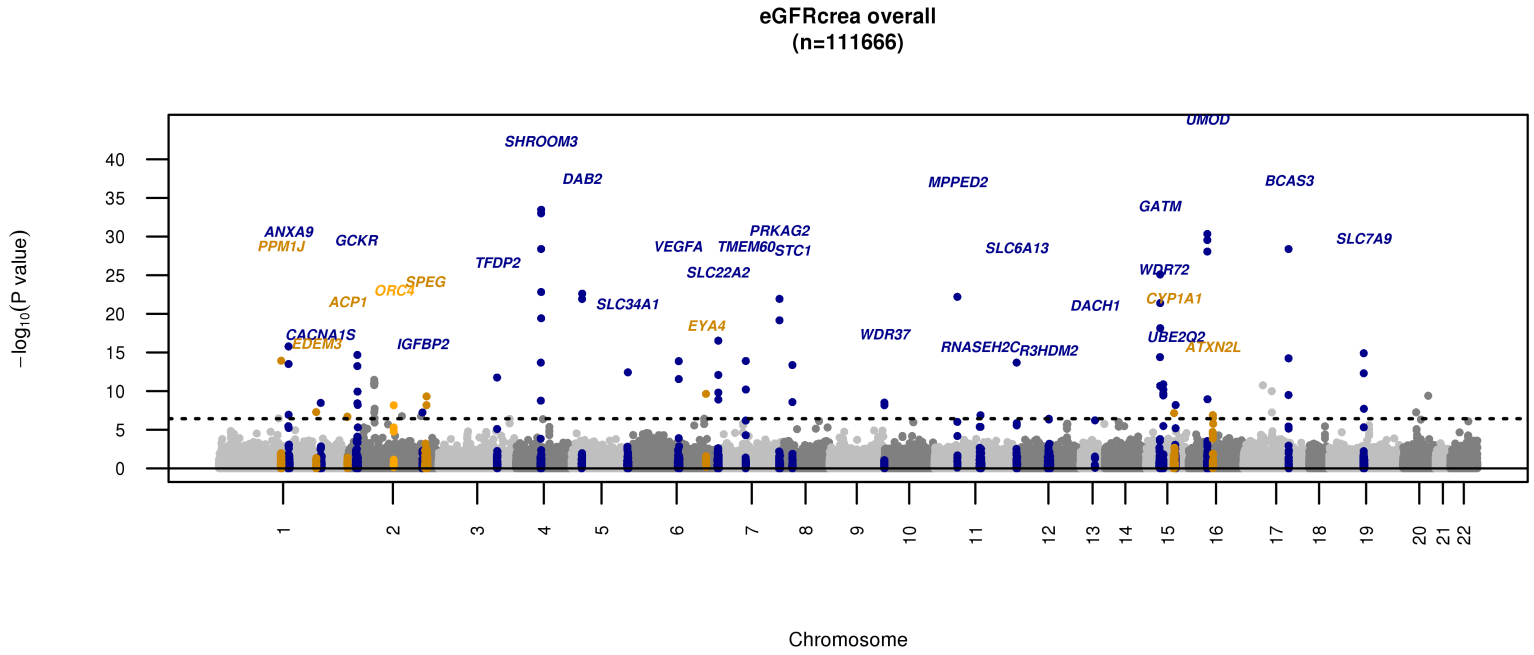
Gene[a]	Chr	cMAF	N Variants[b]	T1[c]			SKAT[d]
				Beta	SE	P	P
LRP2	2	0.070	38	0.003	0.002	6.7E-02	3.5E-7[e]
SLC47A1[f]	17	0.033	4	-0.033	0.004	7.8E-15[e]	3.4E-12[e]
SOS2[g]	14	0.040	8	0.020	0.004	3.3E-06	5.4E-08[e]

[a] Gene name.
[b] Number of variants used in analysis.
[c] The standard burden test collapses the variants with MAF<1% into a single variable and tests the association between this variable with a phenotype[30].
[d] The sequence kernel association test (SKAT) aggregates individual variant score test statistics[32].

[e] Meets chip-wide significant threshold, $P<2.5x10^{-6}$.
[f] Gene-based association results driven by 1 variant.
[g] Novel gene.

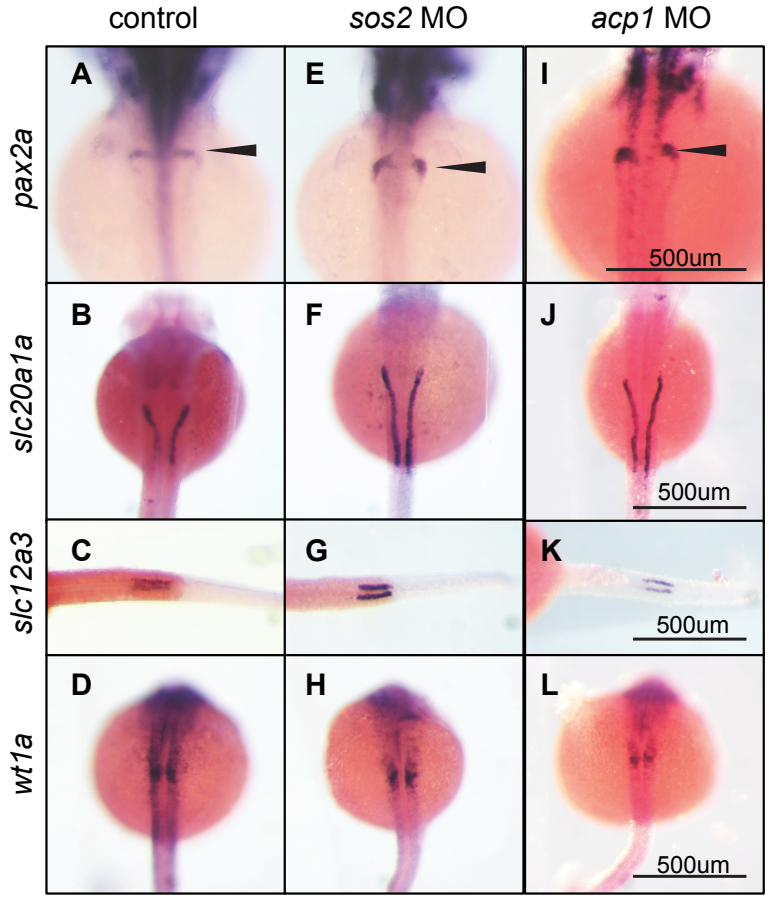
Abbreviations: Chr = chromosome, cMAF = cumulative minor allele frequency used in analysis, SE = standard error, P= p-value

Figure 1.

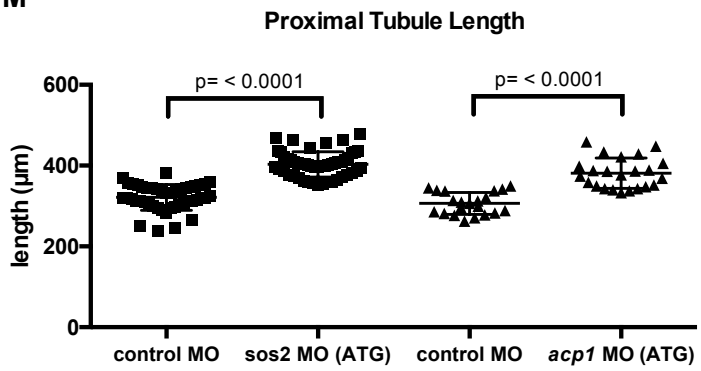

eGFRcrea overall
(n=111666)
−log10(P value)
Chromosome
SHROOM3
DAB2
UMOD
MPPED2
BCAS3
GATM
ANXA9
PPM1J
GCKR
TFDP2
VEGFA
TMEM60
PRKAG2
STC1
SLC6A13
SLC7A9
WDR72
SLC22A2
SPEG
ORC4
ACP1
SLC34A1
DACH1
CYP1A1
EYA4
CACNA1S
EDEM3
IGFBP2
WDR37
RNASEH2C
R3HDM2
UBE2Q2
ATXN2L

42

Figure 2.



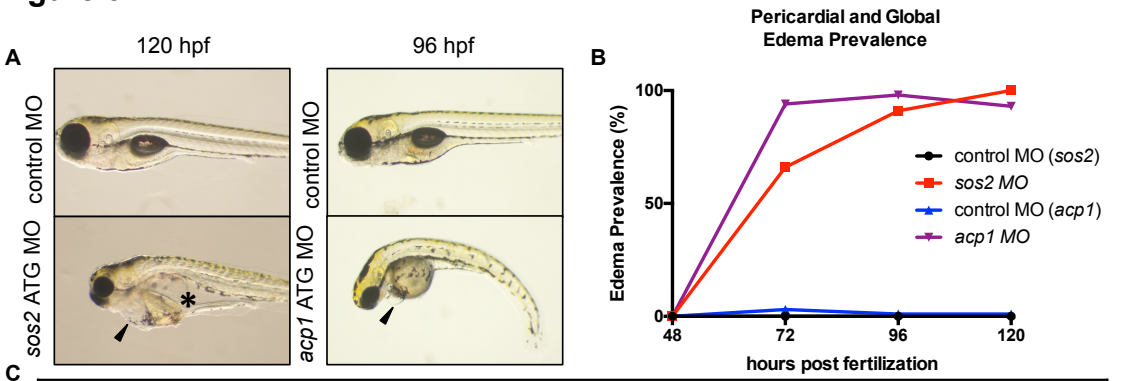
M



N **Number of abnormalities/number examined**

	Global Renal Marker (*pax2a)*	Podocyte Marker (*wt1a)*	Distal Tubular Marker (*slc12a3)*
control M0 (150uM)	0/15 (0%)	0/41 (0%)	0/29 (0%)
sos2 MO (ATG,150uM)	14/19 (74%)	1/47 (2%)	3/44 (7%)
P value	0.00001	1.00	0.27
control MO (300uM)	1/64 (2%)	0/25 (0%)	3/40 (8%)
acp1 MO (ATG, 300uM)	29/38 (76%)	0/39 (0%)	39/40 (98%)
P value	<0.00001	N/A	<0.00001

Figure 3.



C

Edema Prevalence						
Control MO				***sos2* ATG MO**		
	pericardial	global	dead	pericardial	global	dead
48 hpf	0/40 (0%)	0/40 (0%)	0/40 (0%)	0/35 (0%)	0/ 35 (0%)	0/35 (0%)
72 hpf	0/40 (0%)	0/40 (0%)	0/40 (0%)	23/35 (66%)***	0/35 (0%)	0/35 (0%)
96 hpf	0/40 (0%)	0/40 (0%)	0/40 (0%)	32/35 (91%)***	0/35 (0%)	0/35 (0%)
120 hpf	0/40 (0%)	0/40 (0%)	0/40 (0%)	7/35 (20%)**	16/35 (46%)***	12/35 (34%)***

Control MO				***acp1* ATG MO**		
	pericardial	global	dead	pericardial	global	dead
48 hpf	0/79 (0%)	0/79 (0%)	0/79 (0%)	0/101 (0%)	0/101 (0%)	0/101 (0%)
72 - 80 hpf	0/79 (0%)	2/79 (3%)	0/79 (0%)	95/101 (94%)***	0/101 (0%)	2/101 (2%)
96 hpf	0/79 (0%)	1/79 (1%)	2/79 (3%)	97/99 (98%)***	0/99 (0%)	1/99 (1%)
120 hpf	0/77 (0%)	1/77 (1%)	0/77 (0%)	91/98 (93%)***	0/98 (0%)	6/98 (6%)*